

# Identificação de Coesão Social em AVEAs

Helvia P. P. Bastos  
IF Fluminense / PGIE-CINTED  
Dr. Siqueira, 273, Pq. Dom Bosco  
Campos, RJ, Brasil  
helviabastos@yahoo.com.br

Magda Bercht  
PGIE-CINTED, UFRGS  
Porto Alegre, RS, Brasil  
bercht@inf.ufrgs.br

Leandro Krug Wives  
PGIE-CINTED / PGCC  
Instituto de Informática, UFRGS  
Porto Alegre, RS, Brasil  
wives@inf.ufrgs.br

Yasmmin C. Martins  
IF FLuminense  
Dr. Siqueira, 273, Pq. Dom Bosco  
Campos, RJ, Brasil  
nim\_asay@hotmail.com

## ABSTRACT

The study presents a lexicometric analysis of terms of sociability in educational chats. We designated three lexical classes as indicators of social cohesion following the model given by Rourke et al. for assessment of social presence in discursive interactions on line. Data obtained in the processing of occurrences show the viability of identifying the variation of social cohesion expressions, and its possible use as a tool for diagnosing student participation in virtual learning environments. *Key words:* Lexicometry, Social Presence, Cohesion, Chats.

## RESUMO

O estudo apresenta um experimento de análise de termos que indicam a sociabilidade em chats em ambientes virtuais de ensino e aprendizagem. Foram designadas três classes indicadoras de coesão conforme o modelo de Rourke et al. para verificação de presença social em interações on line. Os dados obtidos na análise das postagens demonstram a viabilidade de se identificar as variadas formas de ocorrência de coesão e sua possível utilização como instrumento de diagnóstico da participação do aluno na educação a distância.

Palavras-chave: Lexicometria. Presença Social, Coesão, Chat Educacional.

## 1. INTRODUÇÃO

Este artigo apresenta e discute os resultados obtidos no processamento computacional de sessões de chats em Ambientes Virtuais de Ensino e Aprendizagem (AVEA) em que se buscou verificar a ocorrência de termos lexicais denotadores de coesão, um dos aspectos indicativos de presença social (PS) no modelo de Rourke et al. (1). O conceito de presença social é tido como o grau (*degree*) em que os indivíduos se sentem efetivamente conectados uns aos outros em um evento de comunicação mediada por computador – CMC (2, 3).

Interações on line baseadas em texto podem fomentar as relações de pertencimento – sentimento de ligação ou de participação experimentado pelos membros num determinado grupo (4) – em comunidades de aprendizagem, contribuindo para a motivação e retenção de alunos em AVEAs (5, 6).

Segundo Tu (7), pouca presença social pode comprometer a experiência didática porque implica em menor interação social, aspecto fundador da educação a distância.

O corpus de análise é constituído por seções de chat disponibilizadas em um curso de especialização *lato sensu* a distância. O estudo fundamentou-se na perspectiva pragmática de análise das manifestações em situações reais de uso e no conceito de “registro” de Halliday (8), segundo o qual as escolhas lexicogramaticais do produtor textual variam conforme o “contexto da cultura” e o “contexto da situação” que envolvem a produção dos enunciados. Esses princípios são intimamente relacionados à noção de gênero textual dada por Bakhtin (9) e atualizada para a CMC por Marcuschi (10).

O experimento de processamento automático de chats em AVEAs, parte da pesquisa desenvolvida no programa de doutoramento em Informática na Educação (PGIE - UFRGS), foi feito no programa “Compostas\_V1” (11), desenvolvido no projeto de iniciação científica, “Mineração de dados textuais eletrônicos”, do Núcleo de Informática na Educação do Instituto Federal Fluminense, Campos, RJ. Neste estudo, buscamos organizar as classes de coesão social segundo as diferentes ocorrências encontradas nos chats e estabelecer possíveis comparações com os resultados. Esse trabalho de processamento textual visa ao desenvolvimento de uma ferramenta a ser adicionada a plataformas de educação a distância (em especial, ao ambiente Moodle (12). Pretende-se que ela funcione como recurso auxiliar de tutoria realizando diagnósticos de PS pelo processamento de comunicações por escrito.

## 2. FUNDAMENTOS TEÓRICOS

Esta seção conceitua, brevemente, as noções de “gênero textual chat”, “coesão social” e “lexicometria”.

### 2.1. Chat como gênero textual

A partir de sua acepção da língua como uma atividade social e do princípio de dialogicidade que fundamenta a comunicação humana, Bakhtin (8) afirma que os sujeitos interagem linguisticamente segundo os propósitos comunicativos do evento comunicativo. Dessa forma, os enunciados são estruturados em “gêneros textuais” (também “gêneros do discurso” ou “gêneros discursivos”). Segundo Marcuschi (10), os “gêneros textuais digitais” são formas discursivas híbridas independentes ou mescladas com os convencionais, acrescidas das peculiaridades do suporte digital (p.ex. uso de caracteres alfanuméricos). O gênero “chat” se caracteriza, dentre vários aspectos, pela estruturação freqüente em frases curtas e fragmentadas, por repetições e reconstruções da(s) idéia(s). Marcuschi (13) reitera que a redação

nos chats não configura “uma nova forma de escrita”, mas uma forma diferente de nos relacionarmos com ela.

Os chats são tipificados, também, pela incorporação de traços característicos da fala coloquial oral à escrita digital. Não contando com as pistas fornecidas na comunicação face a face (p.ex. gestos, expressão facial, entonação), os interlocutores usam os recursos gráficos e hipermediáticos da máquina para (re)criar marcas linguísticas e paralinguísticas (p.ex. *emoticons* e *gifs*). Assim, a escrita em ambientes virtuais é pautada, especialmente nos chats, por reduções gráficas (*q* = que; *td* = tudo), onomatopéias (*hehehe* = risos), alongamentos gráficos (*oiiii* = oi), uso de emoticons, substituição da acentuação por consoantes (*eh* = é), uso de maiúsculas para enfatizar idéias ou atrair atenção (*ADOREI* = adorei), entre outros.

## 2.2. Presença Social e Coesão

A noção de Presença Social (PS) usada neste trabalho se apóia na “Teoria do Processamento Social” de Walther (14), segundo a qual tanto face a face quanto via CMC, o conteúdo linguístico das interações é suficiente para que os interlocutores formem impressões sobre os outros. A partir desse pressuposto, Rourke et al. (1) listaram categorias e indicadores de PS (Quadro I) para análise de conversações on line de forma a sinalizar o nível de envolvimento do aluno e de seu sentimento de pertença na experiência de aprendizagem.

**Quadro 1. Categorias de PS na CMC**  
(adaptado de Rourke et al. 2001)

Categorias	Indicadores
Afetividade	expressão da emoção, utilização de humor, auto-revelação
Interatividade	participação contínua na discussão, referência explícita a outras postagens, encorajamento, apreciação e concordância
Coesão	saudações, despedidas, expressões fáticas, vocativos, pronomes inclusivos

Neste artigo, usamos o termo “coesão social” para estabelecer uma distinção com “coesão” ou “coesão textual”, denominações usadas correntemente em Linguística para se referir aos aspectos que dão progressão e continuidade ao texto (15). Em seu trabalho, Rourke et al. (1) explicam que a categoria “coesão” ilustra ações que produzem, fomentam e mantêm o sentimento de compromisso no grupo. Assim, os indicadores de coesão em interações em comunicações baseadas em texto incluem os seguintes recursos discursivos:

(i) uso de termos de sociabilidade e expressões fáticas ou votivas em cumprimentos e despedidas (p. ex. *Oi, Boa noite, tchau, bom feriado!, tô saindo, bons estudos para todos*). Conforme Yetere, Steve e Gibbs (16), expressões fáticas não visam transmitir idéias e fatos, mas fortalecem a proximidade e “confirmam que a comunicação está, de fato, acontecendo”. Essas expressões aparecem no *corpus* de estudo acompanhadas, ou não, por vocativos específicos (nomes próprios ou titulação) ou genéricos (p.ex. *peçoal*).

(ii) endereçamento e referência aos demais usando pronomes pessoais, possessivos (p.ex. *tu, vocês, a gente, nosso, todos nós*), indefinidos (p.ex. *todos*), nomes próprios, apelidos ou outros tipos de vocativo (p.ex. *colegas, Prof.*).

## 2.3. Lexicometria Computacional

McEnery e Wilson (17) definem “lexicometria” como o levantamento estatístico da frequência e variação de palavras num determinado texto. Utilizando recursos do computador, a Lexicometria Computacional é um procedimento utilizado nas áreas de Linguística de Corpus e Descoberta de Conhecimento em Textos, possibilitando análises qualitativas a partir de dados de natureza quantitativa. A extração dos dados lexicais permite identificar diferentes relações entre eles, por exemplo, aspectos de similaridade ou de desvio (18)

Em tarefas de Lexicometria auxiliadas por computador e na Mineração de Textos, costuma-se preparar o *corpus* de análise, extraído termos e outros elementos considerados não necessários ou pouco relevantes. Essa fase do trabalho, denominada de “pré-processamento” é seguida, então, da extração e análise do conhecimento propriamente dito, etapas comentadas nas próximas seções.

## 3. METODOLOGIA

O experimento buscou identificar termos denotadores de Coesão Social segundo as classes discriminadas em 3.3 e a frequência dessas ocorrências nas mensagens do chat observado. O software *Compostas\_v1* foi desenvolvido, inicialmente, para permitir a identificação de termos compostos, comuns em expressões de cumprimento e despedida como *bom dia* e *até logo*.

Outras informações sobre o software usado na pesquisa, o *corpus* de estudo e as etapas do experimento são apresentados a seguir.

### 3.1. Software usado na análise lexicométrica

O experimento foi realizado no *Compostas\_v1* (10) programa de mineração desenvolvido inicialmente para processar termos compostos. Neste trabalho, “termos compostos” são aqueles formados por vocábulos que, juntos, possuem um sentido diferente dos mesmos em separado (ex. *bom dia*).

Para o experimento, foram acrescentadas novas funções ao *Compostas\_v1*: arquivos com listas de palavras indicadoras de coesão verificadas na análise textual convencional, um contador para verificação do número total de palavras e outro para contagem das ocorrências de Coesão Social. Foi criada, também, uma janela auxiliar para inserção de termos pertencentes a cada lista.

### 3.2. Corpus de análise

O *corpus* usado no estudo é constituído por 8.779 entradas de diálogo totalizando 27.624 palavras extraídas de 228 seções de chat realizadas no “Curso de Especialização em Informática Instrumental para Professores da Educação Básica” (19), ofertado para professores da rede estadual em cinco pólos do Rio Grande do Sul. O curso, desenvolvido pelo Instituto de Informática e Secretaria de Educação a Distância (SEAD) da Universidade Federal do Rio Grande do Sul em convênio com a Universidade Aberta do Brasil (UAB), constou de oito disciplinas ofertadas na plataforma Moodle de setembro de 2009 a novembro de 2010 com participação final de cerca de 130 alunos.

### 3.3. Etapas de limpeza e indexação

Como a pesquisa enfoca apenas manifestações discursivas feitas por alunos, foi feita a exclusão manual de postagens realizadas por tutores, resultando em 7.627 entradas para o processamento.

Nessa etapa, extraiu-se, também, todo tipo de imagem e as chamadas “stopwords” (termos de baixa frequência e relevância). As categorias gramaticais excluídas foram: artigos definidos e indefinidos, pronomes relativos e demonstrativos, numerais, preposições, advérbios, e verbos de ligação. A palavra “até” (preposição ou advérbio) não foi excluída porque é frequentemente usada em despedidas (p.ex. *até logo*).

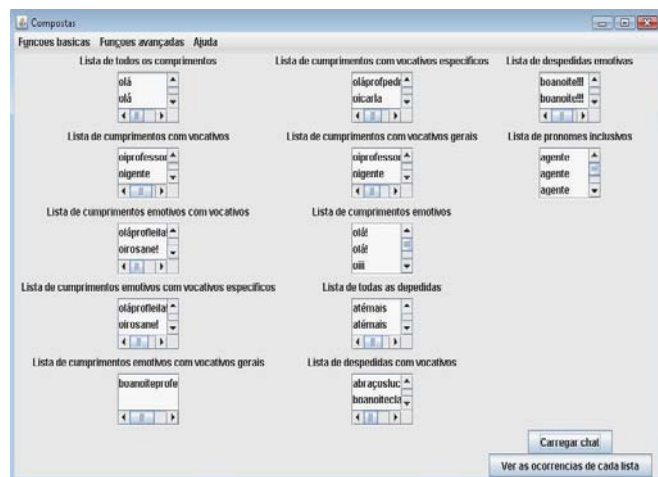
Para que o Compostas\_v1 pudesse reconhecer e computar manifestações de Coesão Social no corpus de análise, foram criadas listas de ocorrências possíveis, organizadas em três classes – *cumprimentos*, *despedidas*, *endereçamento*, que, por seu turno, apresentam categorias gramaticais em comum como vocativos, ponto de exclamação e expressões votivas. Uma vez que a escrita nos chats ocorre em tempo real e recorre a diferentes artifícios para simular a conversa face a face, o experimento aqui descrito desconsiderou os erros de digitação e o uso não convencional de maiúsculas. A lista de classes ficou assim caracterizada:

**a) Saudações** (com exemplos): a.1. termo de saudação: *olá, oi*; a.2. termo de saudação + exclamação: *olá!*; a.3. termo de saudação + vocativo genérico ou específico: *oi gente, olá Prof.*; a.4. termo de saudação + vocativo genérico ou específico + exclamação: *oi meninas!, olá professor!*; a.5. termo de saudação + pronome: *olá para todos(as)*

**b) Despedidas** (com exemplos): b.1. termo simples de despedida: *tchau, abraços, aguardo*; b.2. termo de despedida + exclamação: *Boa Noite!!!, abraços!!*; b.3. expressão votiva: *Boas aulas!, bom feriado*; b.4. termo de despedida + pronome: *abraços a todos*

**c) Pronomes**: tipos de ocorrências e exemplos: c.1. de endereçamento: *tu / você(s) / teu(s) / tua(s) / seu(s) / suas(s) / todos (você(s) / eles)*; c.2. de inclusão: *nós / a gente / nosso(s) / nossa(a) / todos(as) / todos (nós)*

O processamento das classes incluiu a criação de (i) uma lista de palavras – recurso que recebe os termos determinados para processamento; (ii) um contador de palavras – funcionalidade que permite a contagem de mais de uma ocorrência do termo.



**Figura 1. Janela do Compostas\_v1 com as classes e listas de ocorrências de Coesão Social**

## 5. RESULTADOS

Após o trabalho de remoção de *stopwords*, o estudo lexicométrico da “Coesão” possibilitou a identificação e contagem de 27.275 palavras denotadoras de sociabilidade assim distribuídas:

**Quadro 4: Levantamento estatístico de termos de Coesão Social**

Classes	Ocorrências	Percentual
Saudações	136	0.3 %
Despedidas	302	1.04 %
Pronomes	2.100	7.7 %

Os percentuais 0.3% (Saudações) e 1.04% (Despedidas) podem parecer baixos em relação ao total de entradas (7.627) feitas por alunos nas seções analisadas. Verifica-se, entretanto, que essas conversas são constituídas, quase sempre, por longos diálogos iniciados e finalizados com algum termo de coesão, caracterizando o que Kerbrat-Orecchioni (20) classifica como “trocas rituais”. É preciso atentar que tanto nas saudações quanto nos cumprimentos, é comum se iniciar ou se fechar a conversa apenas com vocativos específicos ou genéricos (Cf. 2.2.).

Como os vocativos são formas de chamamento ou interpelação dirigidas ao ouvinte / leitor, eles são usados em diferentes partes da sentença, isto é, não necessariamente na introdução ou fechamento dos diálogos. Assim, o levantamento do uso de vocativos no experimento não contempla as ocorrências totais dessa estratégia discursiva nas postagens analisadas. Os valores apresentados no Quadro 5 são, portanto, *parciais* e *meramente ilustrativos*, uma vez que não foram contabilizados todos os nomes de alunos e tutores, ou seja, apenas o nome de um determinado aluno – Aluno A, e de um determinado tutor – Tutor A.

**Quadro 5: Uso de vocativos no corpus de estudo**

Tipos de Vocativo	Ocorrências
Específico ( <i>Professor/a, Prof., profe</i> )	130
Específico (aluno A)	44
Específico (tutor A)	160
Genérico ( <i>gente, colegas, meninas, pessoal</i> )	85

No que diz respeito aos pronomes, os resultados mostrados no Quadro 4 referem-se às ocorrências totais de pronomes de endereçamento e de inclusão (2.100), quantidade que ilustra como os participantes se posicionam no grupo, em permanente referência aos demais e a eles mesmos como parte de uma coletividade.

Vale atentar para o fato de que, em Português, a função sintática de “sujeito” exercida pelo pronome pessoal inclusivo “nós” é frequentemente marcada pela desinência número-pessoal “-mos”. No texto de análise, aparecem, por exemplo, 12 ocorrências de “podemos”, 9 de “ficamos”, e 8 de “sabemos”, por exemplo. Como esse recurso da língua não foi previsto na preparação do programa, o resultado final para a classe “pronomes” não reflete, exatamente, os vários modos como os participantes manifestam sua inclusão no grupo.

## 6. CONCLUSÕES E ENCAMINHAMENTOS

A criação de novas funcionalidades no software Compostas\_v1 possibilitou identificar e contabilizar termos de sociabilidade indicadores de “Coesão Social” em seções de chat educacional no modelo proposto por Rourke et al. (1). Essas funções precisam ser refinadas de modo a atender à variedade de ocorrências lexicais, à complexidade da expressão discursiva dos produtores textuais, e às particularidades da escrita no gênero chat. Um aspecto não contemplado no experimento foi a identificação de formas verbais na 1ª pessoa do plural sem o pronome inclusivo “nós”, um aspecto da Língua Portuguesa que precisa ser considerado em estudos futuros.

Como resultado positivo obtido no experimento, citamos a possibilidade de se identificar termos compostos e de se fazer procedimentos lexicográficos a partir da elaboração de classes e respectivas listas abrangendo os variados recursos léxicogramaticais da língua disponíveis ao produtor textual..

Considerando que à versão original do software Compostas foram acrescentadas novas funções, o programa passa, doravante, a ser chamado de Compostas\_v2.

Pretende-se, como próxima fase da pesquisa, realizar experimentos semelhantes com as demais categorias indicadoras de Presença Social (PS) não contempladas neste estudo – *afetividade e interatividade*, seja em seções de chat ou em fóruns de discussão disponibilizados em ambientes de educação a distância criados na plataforma Moodle.

## 7. REFERÊNCIAS

- [1] Rourke, A., Anderson, T., Garrison, D.A., Archer, W. 2001. Assessing social presence in asynchronous text-based, computer conference. *Journal of Distance Education*. 14, 2. pp. 50-71.
- [2] Swan, K., Richardson, J. C., Garrison, D.R. 2008. Validating a measurement tool of presence in online Communities of Inquiry. *E-Mentor*. 2, 24. <http://www.e-mentor.edu.pl/artykul/index/numer/24/id/543>
- [3] Na Ubon, A., Kimble, C. 2004. Exploring social presence in asynchronous text-based online learning communities. *5th International Conference on Information Communication Technologies in Education*. Atenas, Grécia, 2004. pp. 292-7.
- [4] Palácios, M. 1995. O medo do vazio: comunicação, sociabilidade e novas tribos. In Rubim, A. A. *Idade Mídia*, Edufba, Salvador, BA.
- [5] Garrison, D. R., Anderson, T. 2003. *E-learning in the 21st century: a framework for research and practice*. RoutledgePalmer, Londres.
- [6] Mackey, K.M., Freyberg, D. L. 2010. The effect of social presence on affective and cognitive learning in an international engineering course taught via distance learning. *Journal of Engineering Education*. 99, 1, pp. 23-34.
- [7] Tu, C. H. 2002. The impacts of text-based CMC on online social presence. *The Journal of Interactive Online Learning*. 1, 1. <http://www.ncolr.org>
- [8] Halliday, M.A.K. 1978, p. 31 Language as social semiotic: the social interpretation of language and meaning. Edward Arnold, Londres. p. 31.
- [9] Bakhtin, M. 2000. *A estética da criação verbal*. 2.ed. Huicitec, São Paulo.
- [10] Marcuschi, L.A. 2004. Gêneros textuais emergentes no contexto da tecnologia digital. In Marcuschi, L. A., Xavier, A.C. (orgs.). *Hipertexto e gêneros digitais: novas formas de construção de sentido*. Lucerna, Rio de Janeiro, pp. 13-67.
- [11] Martins, Y., Azevedo, B., Bastos, H. 2010. Identificação de palavras compostas como auxílio à Mineração de Textos: desenvolvimento do Compostas\_v1. *VI Circuito de Tecnologias da Informação*. Campos dos Goytacazes, RJ, 2010.
- [12] Moodle – Course Management System. 2011. <http://www.moodle.org/>
- [13] Marcuschi, L. A. 2001. Fala e escrita: uma visão não-dicotômica. *Revista do GELNE*, 3, 1 jan-jul 2001. pp. 61-7.
- [14] Walther, J.B. 1992. Interpersonal effects in computer-mediated interaction: relational perspective. *Communication Research*. 19, 1. pp. 52-90.
- [15] Adams, J. M. A. 2008. *A linguística textual: introdução à análise textual dos discursos*. Cortez, São Paulo. p.60.
- [16] Vetere, F., Howard, S., Gibbs, M. 2005. Phatic technologies: sustaining sociability through ubiquitous computing. *Conference in Human Factors in Computing Systems CHI 2005*, Portland, EUA, Abril 02-07. (Workshop). <http://www.vs.inf.ethz.ch/events/ubisoc2005/UbiSoc%202005%20submissions/12-Vetere-Frank.pdf>
- [17] McEnery, W., Wilson, A. 2004. *Corpus linguistics: an introduction*. 2. ed. Edinburgh University Press, Edinburgh, RU.
- [18] Wives, L., Rodrigues, N.A. 2000. *Eurekha*. *Revista Eletrônica da Escola de Administração da UFRGS – Read*, 17, 6, set/out. 2000.
- [19] URL: <http://www.inf.ufrgs.br/informatica-uab>
- [20] Kerbrat-Orecchioni, C. 2006. *Análise da conversação: princípios e métodos*. Parábola, São Paulo.